東北大学機関リポジトリ
TOUR
Tohoku University Repository
東北大学

Polygraph Techniques: The Influence of the Stimulation Test

著者	TONNEY Jaime, YAMAMOTO Naohiro, HATAYAMA Toshiteru
journal or publication title	Tohoku psychologica folia
volume	62
page range	99-108
year	2004-04-01
URL	http://hdl.handle.net/10097/54741

Tohoku Psychologica Folia
2003, **62**, 99－108

Polygraph Techniques: The Influence of the Stimulation Test

JAIME THONNEY[1], YAMAMOTO NAOHIRO (山本直宏)[1],
and HATAYAMA TOSHITERU (畑山俊輝)[1]
(*Tohoku University*)

The influence of the 'stimulation test', part of the pre-test of a polygraphic interrogation in many places, was studied as a way to address the issue of the effect of subjects' beliefs on the technique's precision. A second objective of this study was to investigate the differential accuracy provided by three different parameters (heart rate and period, skin response) when used individually. Twenty university students participated in an experiment that used the stimulation test as the manipulated variable; ninety-nine trials were completed. The hypothesis (that the accuracy of the polygraph would increase if the expectations of the subjects were manipulated, as accounts from the field practice seem to indicate) could not be substantiated. The single individual parameter that provided better results seemed to be the galvanic skin response, which coincides with many experimental reports.

Key words: polygraph, lie detection, stimulation test, autonomic responses

Introduction

Polygraphic interrogation is an area of applied psychology because the lie detector is a psychological test based on assumptions about the relationships between emotional and cognitive states on one hand and physiological responses on the other; therefore its study provides the opportunity to address issues connected to human emotion, and its psycho-physiological correlates. This research started through a theoretical study about the unsolved methodological and theoretical problems of the phenomenon of lie detection through the polygraph; a second step consisted in the design of an experimental approach that addressed them. Specifically, the following pending issues were identified: (1) discrepancies between data provided by experimental research and the practice of lie detection; (2) problems concerning the process of scoring the obtained data; difference of evaluation criteria between laboratory research and field work; lack of clearly defined and universally accepted criteria; (3) problems concerning the difficulty of reproducing in an experimental context some of the variables present in real-life situations; (4) problems concerning the examinees: the credibility of the polygraph and its repercussions.

Objectives of the Experiment (i):

To investigate the influence of subjects' expectations about the accuracy of the test and its final effectiveness. To do so, the *stimulation test*, part of the pre-test of a polygraph examination,

1. Department of Psychology, Graduate School of Arts and Letters, Tohoku University, Kawauchi, Aoba, Sendai 980-8576, Japan

was used; The stimulation test is supposed to increase the subjects' expectations about the accuracy of the polygraph; it is presented by many authors as a way to 'reassure the innocent suspect and increase the fear of detection of the guilty one' (Matte, 1996). It usually consists of a manipulation of the suspects' beliefs about the polygraph through a deception from the polygraphers, who use marked cards (Orne and Waid 1981), or use a deck that contains only one type of card (Vrij, 2000). According to our hypothesis, the polygraph test would render higher accuracy rates when such manipulation was applied.

Objectives of the Experiment (ii):

To explore the correlation between the different parameters used to evaluate the polygraph technique charts, and the accuracy they provide.

Method

Participants

Twenty subjects between 20 and 30 years old took part in the experiment. Seventeen of them were male and 3 female; all of them were Psychology students at Tohoku University; ninety-nine trials in total were completed.

Apparatuses

Skin conductance response meter 'BioDerm' model 2701; perspiration meter SKD-2000; breathing pick-up 'Nihon Kohden' SR-601S; Biotop 6R 12 series amplifier for bio-electrical phenomena; Sony data recorder PC 208 Ax; Windows 'PowerLab' System; Macintosh computer; electrically-shielded room.

Stimuli

We used the same stimuli used in the real-life practice of lie-detection during the stimulation test: cards; the same manipulation used in real-life situation was used as well.

Procedure

Condition I -subjects' expectations not manipulated; no stimulation test applied. The first group -10 participants, 50 trials- was subject to the lie-detection technique without any prior manipulation of their beliefs.

Each subject had to pick one of five cards he could not see; five different cards were used as stimuli, and were shown to the subjects beforehand. Figure 1 shows the five cards as the subjects see them (facing the table) and as the subjects imagine them.

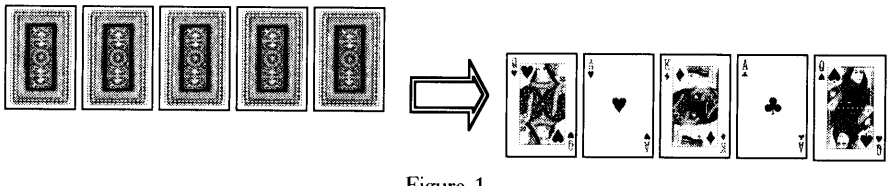

Figure 1

However, the first stimulus was changed (*dummy or buffer* question) since a higher psycho-physiologic response is always expected during the first stimulus of a set; Figure 2 shows the five cards as the subjects see them (facing the table) and as they really are.

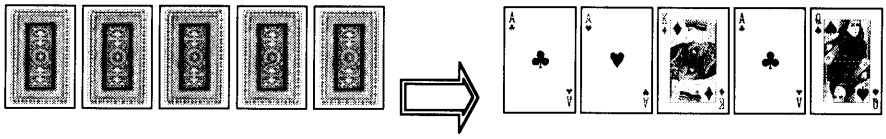

Figure 2

The subjects were asked to reply with one word ('No') to the questions asked by the experimenter outside the shielded room through the intercom (D. Lykken's Guilty Knowledge Technique):

Did you pick the Queen of Hearts? – 'No' -
Did you pick the Ace of Hearts? – 'No'
Did you pick the Ace of Clubs? – 'No'
Did you pick the King of Diamonds? – 'No'
Did you pick the Queen of Spades? – 'No'

The rationales of the technique are that: (1) the psychophysiological indexes change when the subject recognizes a meaningful stimulus; the guilty knowledge detection technique is 'intended to determine whether the subject is aware of certain information: it is he demonstration of such awareness... what might differentiate between a guilty and an innocent suspect.' (Lykken, 1974). Thus, in Japan, where it is usually called 'concealed information test', is mostly regarded as a kind of recognition test and examiners use it to determine if the suspect of committing a crime identifies a crime-related detail (Nakayama, 2002). (2) The highest autonomic response indicates a deceptive statement. According to Lykken: 'Whether [the subject] is high or low in reactivity, we can still expect that his response to this significant alternative will be stronger than to the other -non significant- alternatives.'

The second advantage of the guilty knowledge method is related to the process of habituation; unless a complete habituation has occurred (that is no responses are elicited by any of the stimuli), guilty suspects may show relatively higher responses to the relevant stimuli 'even when both types of responses were attenuated.' (Ben-Shakhar and Elaad, 1997).

The stimuli were shown and the subjects' utterances and psycho-physiological responses were recorded on the hard disk in the PC.

Condition II -subjects' expectations manipulated; stimulation test applied. Each subject had to pick one of five cards he could not see; he believed all the cards were different. Figure 3 shows the five cards as the subjects see them (facing the table) and as the subjects imagine them.

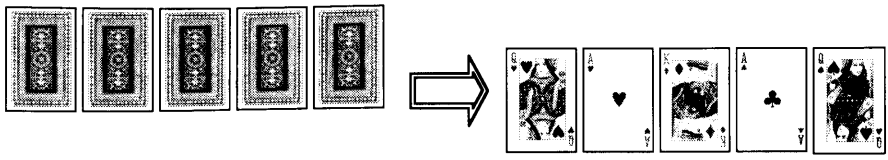

Figure 3

However, following the standard Stimulation Test procedure, they were all the same. During the stimulation test the experimenter always knew which card the subjects had picked for each one of the trial; in fact for this second group the stimuli were both the cards and the experimenters' feedback. Figure 4 shows the five cards as they really are.

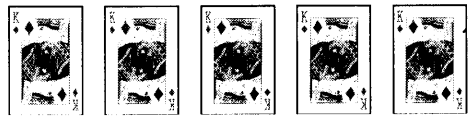

Figure 4

The second group -10 participants, 49 valid trials - was subject to a manipulation of their beliefs about the polygraph through a deceitful feedback from the experimenters. After the three first trials we told the subjects we would analyze their responses and then we 'informed' the subject which response of the five alternatives had been a lie:

'We believe the card you just picked is the King of Diamonds'

In fact all five cards were Kings of Diamonds; then three to five more trials were conducted -this time without any deception. The responses of these new trials were analyzed later to see if the accuracy of the polygraph had increased after we tried to deceive the subjects. The experiment was conducted in a way as similar as possible to a real-life use of the polygraph: the experimenter was 'blind' to the card the subjects picked; the responses were written down on the subjects' files and sealed to be contrasted with the experimenter's scoring of the charts later. All the subjects were deceived with respect to the real mechanism and objective of the experiment; they were not told their expectations would be manipulated. The cards were changed after each trial to avoid suspicions. When all the trials were completed, the accuracy of the polygraph test was compared for the two groups (Condition I & II) according to the three used evaluation parameters.

Data evaluation: criteria

Since the lack of universally accepted criteria for the scoring of the charts resulting of a polygraphic interrogation was one of the conclusions of the theoretical research that provided the basis for this experimental design, clearly defined criteria were used, and they were always of an exclusively quantitative nature. The first of those criteria was an operational definition, elaborated

in virtue of the interest the polygraph community has been showing in the heart response as autonomic index: the heart rate. The operational definition was structured using traditional psychophysiological definitions (Schneiderman and Dauth, 1974). Fifteen heart periods (from systolic tip to systolic tip) after each question were measured; in a second step the minimum heart period was subtracted from the maximum one, and the largest difference was assumed indicative of deception. Figure 5 shows an example of a real subject's polygraph chart and the periods considered after each stimulus.

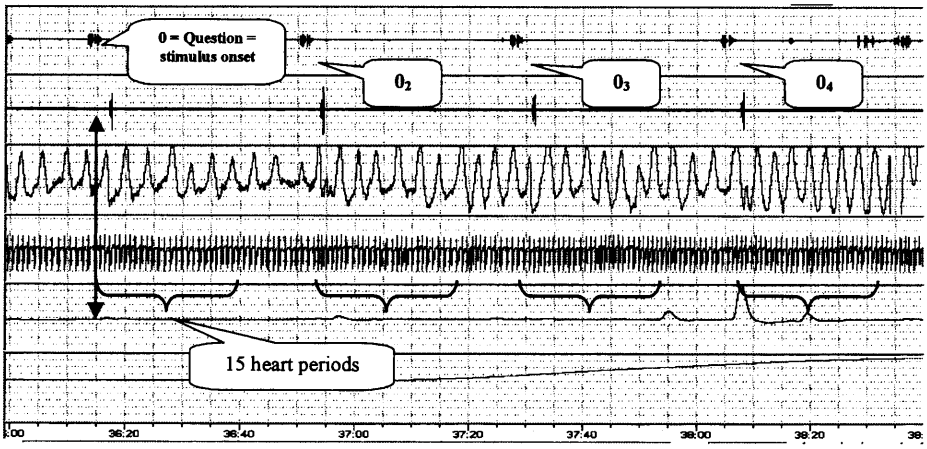

Figure 5. Psycho-physiologic responses: four questions of a trial as seen on the display of the Windows Power Lab system. From top: experimenter's voice (each utterance is a question), subject's voice (each utterance is an answer), breathing, ECG, galvanic skin resistance, perspiration.

The heart rate, a parameter used by professional polygraphers, was used as a second criterion to tell between truthful and deceptive statements. The heart rate is defined as 'the number of heart beats in one minute'; since in a Guilty Knowledge Technique the questions are presented to the subject every twenty seconds (standard real-life practice) a formula was used to calculate it: heart rate = 60 seconds / heart period (in seconds). Once more, the response that presented the highest heart rate of the four alternatives was considered the most likely to be deceptive.

The 3rd evaluation criterion used in this experiment was a combined GSR amplitude and duration measurement, also a real-life practice (Matte, 1996). In principle, the height (amplitude) of GSR responses was considered; however, on some occasions the GSR reactions go up to the limit of the apparatus' sensitivity and remain there for several seconds before returning to baseline (overshooting), making the use of GSR amplitude impossible. When the GSR reaction surpassed the sensitivity level of the apparatus the duration replaced the height of the response as the used criterion, measuring it at overshooting level to equalize the criteria for all four options in each trial.

Sometimes a secondary surge of the GSR pen, known as a 'complex response', may be caused by the examinee's realization and fear that they may have reacted to that particular

question thereby causing a second stimulation of the sweat glands (Fowles, 1974; also Matte, 1996). Since this secondary heave of the GSR pen is supposedly caused by the subject's fear that he may have been detected when lying, both peaks were considered as part of one response and their duration was measured together as long as the time between them was no longer than the duration of any of the responses themselves (operational definition). Figure 6 shows the two aforementioned cases of 'complex response'.

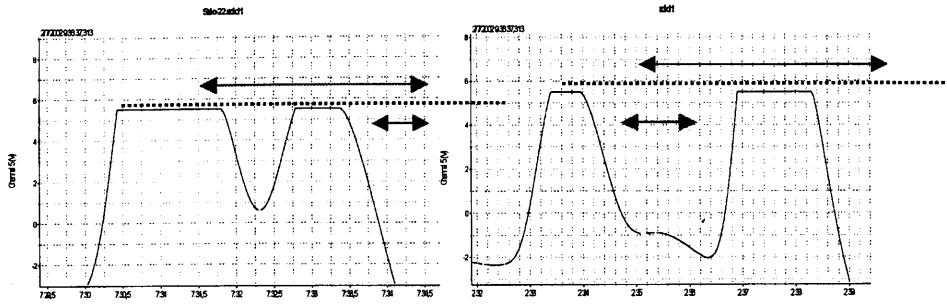

Figure 6. GSR (right): only one response (long line), duration/height of both peaks measured together (overshooting level); GSR (left): two responses, duration/height of only one peak length measured (short line: the first one within the considered period).

In the final step the numerical scores of the HR, HP and GSR responses of each subject were written in a table. The highest scores within each trial were considered the deceptive ones; then the subjects' files were open to contrast those results with the ground truth (the cards the subjects had picked) and determine the accuracy of each parameter. Table 1 shows an example from a real subject; the intensity of the autonomic disturbances was the criterion to determine which utterance corresponded to the deceptive statement.

Table 1 Sample of the experiment's charts; Question 2: Ace of Hearts; Question 3: Ace of Clubs; Question 4: King of Diamonds; Question 5: Queen of Spades.

	GSR amplitude	GSR length	
Question 2	-0,041 V		
Question 3	5,508 V	2,725 sec	Possible deception
Question 4	1,438 V		
Question 5	5,508 V	1,312 sec	

Trial #2	Question # 3 the card should be the Ace of Clubs	Result: SUCCESSFUL DETECTION

The response to the question #3 has been selected as the possible deceitful utterance because the subject's GSR showed the largest duration within the trial. In this case the criterion proved to be right.

Statistical analysis

The ANOVA test was applied to determine the statistical relevance of the differential accuracy of the three parameters considered: heart rate, heart period and galvanic skin response (the variable 'expectations' manipulation' was not considered). The Mann-Whitney test was run to compare the four groups considered and the three parameters used.

Results

The results of the ANOVA test showed no results smaller than 0,05: the differences among them were not significant. The Mann-Whitney test did not show results smaller than 0,05: the differences among the accuracies of the groups were not significant under any parameter. Table 2 shows the accuracy percentages according to condition and autonomic response measured.

Table 2 Accuracy percentages of the GKT (Guilty Knowledge Test) such as applied in this experiment; the accuracy expected according to chance is ¼, that is 25%.

	No stimulation test	Stimulation test
Heart period	**16 %**	**27 %**
Heart rate	**18 %**	**22,44 %**
Galvanic skin response	**32 %**	**30,61 %**

The main hypothesis (that the accuracy of the polygraph would increase if the expectations of the subjects were manipulated, as accounts from the field practice seem to indicate) could not be substantiated. The single individual parameter that provided better results –though not significant from a statistic viewpoint– was the GSR, which coincides with many laboratory reports. The lowest accuracy obtained (16 %) corresponded to heart period as unique parameter to analyze the charts; the highest accuracy obtained (32 %) was provided by the scoring of skin responses (also unique parameter to analyze the charts). Figure 7 shows the obtained accuracy rates in a graph; the most stable autonomic response was the one that provided better accuracy rates: the galvanic skin response.

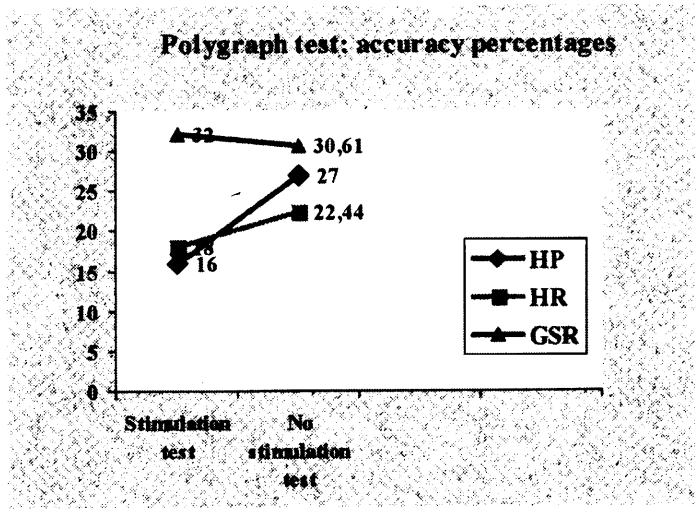

Figure 7. Accuracy percentages.

Discussion

This experiment explored the influence of a real-life practice (the stimulation test) on the results of the polygraph examination. The stimulation test supposedly affects the expectations of the suspects who take the polygraph technique, that is, the experiment addressed the effects of the polygraph's credibility on the outcomes of the technique. Inasmuch as the hypothesis could not be proven, it cannot be said the before-mentioned relationship exists, though it cannot be denied either. What is more, had the results proven such relationship, a valid assumption would have been that the stimulation test had influenced the subjects' expectations and those expectations had affected the accuracy of the test. Since the hypothesis could not be statistically substantiated, it is impossible to determine if the credibility of the test does not have an effect on the polygraph outcomes or not, or if the stimulation test did not affect the subjects' expectations in the first place. It should be noted, though, that the use of a tricked deck of cards is an common practice in the field, albeit only in the pre-test, and to address the validity of its use was the specific objective of this research.

It is important to stress that detection of deception in the field usually involves subject matter that is inherently arousing. The fact that the stimuli used did not have any kind of emotional value may be related to the low accuracy percentages obtained, although once more, it is worth stressing that the procedure followed step-by-step the real-life practice, and the same stimuli were used. Further research should explore the significance of stimuli of different nature in the lie-detection context, and of the link between the subjects' previous exposure to the stimuli. Research conducted in Japan (Nakayama, 2002) support the idea of its relevance.

The validity of the polygraph test in itself was not the issue addressed. Since the experiment

intended to investigate the relative influence of a variable (subjects' expectations), it is valid to determine their comparative influence by using single parameters (ECG period, ECG rate, GSR) and contrasting the results after altering the variables; it is the difference in the accuracies that matters, not the final precision of the test.

In real-life situations most polygraphers have traditionally used a qualitative analysis based on many parameters at the same time; professional polygraphers have specific training and clinical experience. Those may very well be part of the reasons why the accuracy rates found in the literature are many times higher than those obtained in the present research, where an unidimensional analysis at a time was made. The approach of the present experiment has been the use of explicitly defined criteria and exclusively quantitative parameters; such approach was a consequence of the theoretical research done before the experiments, which showed the lack of well-defined scoring criteria accepted by everybody. Thus, clear definitions were provided; such definitions tried to comprise as many elements of the psychophysiological indexes as was possible -as long as they could be quantifiable.

The rationale of many polygraphic techniques states that the larger an autonomic response is, the more likely it indicates a deceitful statements. It is the rationale that best suits experimental research since it allows the experimenters to use of a strictly quantitative analysis of responses under well-defined criteria; it is also the one that best matches the psychophysiological theory of arousal. Although the aim of these experiments was not to investigate it, it is worth noting that such rationale could not be substantiated for all psychophysiological responses. For some autonomic indexes (like breathing, not used in this research, precisely) it does not apply in the least.

The single individual parameter that provided better results seemed to be the galvanic skin response. Many professional polygraphers claim the GSR to be helpful in the laboratory, but of much less value than respiration and heart response in criminal investigations; they seem to believe that the GSR is too responsive to any stimulus. Yet, most of the laboratory studies that have compared GSR with one or more additional variables agree that the GSR is superior to other variables in the detection of deception (Matte, 1996), corroborating this experiment's findings.

It is also important to notice that the participants of this experiment were in all cases subjects with knowledge (both theoretical and practical) of psychology; in many cases specifically of experimental psychology. Some of them were even experienced as researchers and had conducted experiments themselves -albeit not on the topic of lie-detection. The polygraph test was applied to the subjects by people they knew well (their classmates), with whom they interacted in a friendly way in a familiar context (the laboratory of the Psychology Department of Tohoku University), both conditions that hardly enhanced their suggestibility -and the stimulation test is supposed to act on the subjects' suggestibility precisely. These factors let us wonder to what extent their responsivity was not decreased in the first place, due to a previous habituation to the experimental situation. It is worth noting that in some cases (for example the results for Condition I, using heart period as parameter) the accuracy of the test was even lower than the accuracy expected from sheer chance. Some of the traditional criticisms to the polygraph situation in laboratory contexts are valid here. The researchers aimed to bridge the gap between experimental research and field

practice, and therefore used scoring techniques from the latter, though with the strict quantitative viewpoint from the former; yet some of the characteristics of laboratory settings, like the use of a limited and biased sample of participants, could not be prevented. One question that remains and opens new possibilities of research is to what extent the personal traits of the subjects and their previous experience affect the effectiveness of the stimulation test and of the polygraph technique in itself.

References

Elaad, E., & Ben-Shakhar, G. (1997). Effects of item repetitions and variations on the efficiency of the guilty knowledge test. *Psychophysiology* **34(5)**, 587-596.

Fowles, D.C. (1974). Mechanisms of Electrodermal Activity. In Thompson R. F. & Patterson M. M. (Eds.), *Bioelectric Recording Techniques* (pp. 231-271). New York: Academic Press.

Lykken, D. T. (1960). The validity of the guilty knowledge technique: the effects of faking. *Journal or Applied Psychology*, **44**, 258-262.

Lykken, D. T. (1974). Psychology and the lie detector industry. *American Psychologist*, **29(10)**, 725-739.

Matte, J. A. (1996). *Forensic Psychophysiology Using the Polygraph - Scientific Truth Verification - Lie Detection* [CD-ROM]. New York: J.A.M. Publications.

Nakayama, M. (2002). Practical use of the concealed information test for criminal investigation in Japan. In Kleiner, M. (ed), *Handbook of polygraph testing* (pp. 49-86). London: Academic Press.

Orne, M. T., & Waid, W. M. (1981). Cognitive, social, and personality processes in the physiological detection of deception. *Advances in experimental social psychology* Vol. 14.

Schneiderman, N. & Dauth, G. W. (1974). Elechtrocardiogram: Techniques and Analysis. In Thompson R. F. & Patterson M. M. (Eds.), *Bioelectric Recording Techniques* (pp. 165-200). New York: Academic Press.

Vrij, A. (2000). Professional Detection of Deception -The Polygraph. *Detecting Lies and Deceit -The Psychology of Lying and the Implications for Professional Practice* (pp. 169-205). Chichester, UK: Wiley Series in the Psychology of Crime, Policing and Law, John Wiley & Sons Ltd.

(Received January 14, 2004)
(Accepted February 10, 2004)